HHS Public Access

Author manuscript

Cell. Author manuscript; available in PMC 2014 May 23.

Published in final edited form as:

Cell. 2013 May 23; 153(5): 1149–1163. doi:10.1016/j.cell.2013.04.037.

Transcriptional and Epigenetic Dynamics During Specification of Human Embryonic Stem Cells

Casey A. Gifford[1,2,3,*], **Michael J. Ziller**[1,2,3,*], **Hongcang Gu**[1], **Cole Trapnell**[1,3], **Julie Donaghey**[1,2,3], **Alexander Tsankov**[1,2,3], **Alex K. Shalek**[6], **David R. Kelley**[1,3], **Alexander A. Shishkin**[1], **Robbyn Issner**[1], **Xiaolan Zhang**[1], **Michael Coyne**[1], **Jennifer L. Fostel**[1], **Laurie Holmes**[1], **Jim Meldrim**[1], **Mitchell Guttman**[1], **Charles Epstein**[1], **Hongkun Park**[6], **Oliver Kohlbacher**[4], **John Rinn**[1,3,7], **Andreas Gnirke**[1], **Eric S. Lander**[1,8], **Bradley E. Bernstein**[1,5], and **Alexander Meissner**[1,2,3]

[1]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

[2]Harvard Stem Cell Institute, Cambridge, MA 02138, USA

[3]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA

[4]Applied Bioinformatics, Center for Bioinformatics and Quantitative Biology Center, University of Tübingen, 72076 Tübingen, Germany

[5]Howard Hughes Medical Institute and Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA

[6]Department of Chemistry, Harvard University, Cambridge, MA 02138, USA

[7]Beth Israel Deaconess Hospital, Boston, MA 02115, USA

[8]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Abstract

Differentiation of human embryonic stem cells (hESCs) provides a unique opportunity to study the regulatory mechanisms that facilitate cellular transitions in a human context. To that end, we performed comprehensive transcriptional and epigenetic profiling of populations derived through directed differentiation of hESCs representing each of the three embryonic germ layers. Integration of whole genome bisulfite sequencing, chromatin immunoprecipitation-sequencing and RNA-sequencing reveals unique events associated with specification towards each lineage. Dynamic alterations in DNA methylation and H3K4me1 are evident at putative distal regulatory elements bound by pluripotency factors or activated in specific lineages. In addition, we identified germ layer-specific H3K27me3 enrichment at sites exhibiting high DNA methylation in the undifferentiated state. A better understanding of these initial specification events will facilitate

Correspondence: alexander_meissner@harvard.edu.

*equal contribution

identification of deficiencies in current approaches leading to more faithful differentiation strategies as well as provide insights into the rewiring of human regulatory programs during cellular transitions.

Introduction

Coordinated changes to the epigenome are essential for lineage specification and maintenance of cellular identity. DNA methylation (DNAme) and certain histone modifications critically contribute to epigenetic maintenance of chromatin structures and gene expression programs (Zhou *et al.*, 2011; Smith and Meissner, 2013). Genetic deletion of histone methyltransferases and the catalytically active DNA methyltransferases are embryonic or postnatal lethal (Li, 2002) providing evidence for their essential role in proper execution of developmental programs.

Several groups have reported genome-wide maps of chromatin and DNA methylation in pluripotent and differentiated cell types. From these efforts, a global picture of the architecture and regulatory dynamics is beginning to emerge. For example, active promoters generally contain modifications such as H3K4me3 and H3K27ac, while active enhancers are generally enriched for H3K4me1 and H3K27ac (Heintzman *et al.*, 2009; Creyghton *et al.*, 2010; Ernst *et al.*, 2011; Rada-Iglesias *et al.*, 2011). Repressed loci exhibit enrichment for H3K27me3, H3K9me2/3, DNAme, or a combination of the latter two modifications. The enrichment of repressive histone modifications, such as H3K27me3, which is initiated at CpG islands (CGI), is considered a facultative state of repression, while DNAme is generally considered a more stable form of epigenetic silencing (Smith and Meissner, 2013).

Recent studies have reported dynamics that suggest epigenetic priming such as the appearance of euchromatic histone modifications prior to gene activation during *in vitro* T cell differentiation (Zhang *et al.*, 2012) and cardiac differentiation (Wamstad *et al.*, 2012). These results are reminiscent of changes that occur during the early stages of reprogramming towards the induced pluripotent state (Koche *et al.*, 2011) and highlight possible similarities between differentiation and de-differentiation. In parallel to these advances, whole genome bisulfite sequencing (WGBS) has been used to map DNAme genome-wide. Examination of WGBS data from murine ESCs (mESCs) and neural progenitor cells highlighted lowly methylated regions (LMRs) at distal sites that frequently overlap with DNAse I hypersensitive sites (HS) and/or displayed an enhancer signature defined by H3K4me1 and p300 enrichment (Stadler *et al.*, 2011).

Studying the role of epigenetic modifications in the dynamic rewiring of human transcriptional programs *in vivo* is complicated by numerous technical and ethical limitations. However, models for *in vitro* differentiation of hESCs offer a unique opportunity to explore and characterize critical events that prepare, guide and possibly regulate cell fate decisions. Populations representing each embryonic germ layer have been produced from hESCs (Kriks *et al.*, 2011; Chen *et al.*, 2012; Wei *et al.*, 2012).

To dissect the early transcriptional and epigenetic events during hESC specification, we used 2-dimensional, directed differentiation of hESCs to produce representative populations

from the three germ layers, namely ectoderm, mesoderm and endoderm (Hay *et al.*, 2008; Evseenko *et al.*, 2010; Lee *et al.*, 2010) followed by fluorescence-activated cell sorting (FACS) to enrich for the desired differentiated populations. These three cell types, in addition to undifferentiated hESCs (HUES64), were then subjected to ChIP-Seq for 6 histone marks (H3K4me1, H3K4me3, H3K27me3, H3K27ac, H3K36me3 and H3K9me3), WGBS and RNA-Sequencing (RNA-Seq). To complement this data, we also performed ChIP-Seq for three TFs (OCT4, SOX2, NANOG) in the undifferentiated hESCs, as well as ChIP-BS-Seq for FOXA2 in the endoderm population. The combined data sets provide a wealth of information, including holistic views of transcriptional and epigenetic dynamics that help further dissect the molecular events during human germ layer specification.

Results

High Resolution Transcriptional Measurements during Directed Differentiation of hESCs

To better understand the molecular dynamics involved in hESC differentiation, we produced populations representative of each embryonic germ layer, namely ectoderm (Lee *et al.*, 2010), mesoderm (Evseenko *et al.*, 2010) and endoderm (Hay *et al.*, 2008) (see Extended Experimental Procedures). We chose the male hESC line HUES64, an NIH-approved line that readily differentiates into each of the three germ layers. These hESCs can be differentiated into a neuroectoderm-like progenitor population positive for SOX2 and PAX6 by inhibition of TGFβ, WNT and BMP signaling (Figure 1A, **top**). Alternatively, canonical mesoderm markers, such as GATA2 (Figure 1A, **middle**), can be induced using ACTIVIN A, BMP4, VEGF and FGF2 treatment. Lastly, differentiation towards a definitive endoderm-like fate, positive for markers such as SOX17 and FOXA2 (Figure 1A, **bottom**), is induced using ACTIVIN A and WNT3A.

We began by measuring the expression of 541 selected genes, including many developmental transcription factors and lineage markers (Bock *et al.*, 2011), at 24-hour intervals during differentiation towards each respective germ layer. We found that 268 of these genes exhibit expression changes (z-score $\log_2$ expression) during the first five days of differentiation (Figure 1B). Mesendodermal genes, such as *EOMES*, *T*, *FOXA2* and *GSC*, are upregulated at 24 hours of mesoderm and endoderm induction, but not ectoderm differentiation (Figure 1B–C, Table S1). *GSC* expression decreases within 48 hours of differentiation in the mesoderm-like population, while the expression level is maintained in the endoderm population (Figure 1B–C). *EOMES* and *FOXA2* expression is also maintained in the endoderm population accompanied by upregulation of *GATA6*, *SOX17* and *HHEX* (Figure 1B). After transient upregulation of mesendodermal markers, activation of mesodermal markers such as *GATA2*, *HAND2*, *SOX9* and *TAL1* is detected specifically in the mesoderm conditions (Figure 1B–C, Table S1). None of these markers are detected during early ectoderm differentiation, which instead upregulates neural markers, such as *PAX6*, *SOX10* and *EN1* (Figures 1B–C, Table S1).

We found that *POU5F1* (OCT4), *NANOG* and to some extent *SOX2* expression is maintained in our endoderm population (Figure 1B, D, Table S1). This is consistent with prior studies indicating that OCT4 and NANOG expression is detected during the course of early endoderm differentiation and supports NANOG's suggested role in endoderm

specification (Teo et al., 2011). *SOX2* expression is downregulated in mesoderm and to a lesser degree in endoderm, but maintained at high levels in the ectoderm population (log_2 expression 10.9) (Figure 1D, Table S1), while *ZFP42* (REX1) is similarly downregulated in all three lineages (Figure 1B, Table S1). We confirmed that these populations indeed represent a precursor stage for each respective lineage by inducing them to differentiate further, which resulted in upregulation of genes such as *OLIG2* and *SST* in the ectoderm (Chambers *et al.*, 2012), *TRPV6* in the mesoderm (Evseenko *et al.*, 2010), and *AFP* and *HGF* in the later endoderm populations (Figure S1A) (DeLaForest *et al.*, 2011). Lastly, multidimensional scaling confirmed that at 24 hours the mesoderm population is very similar to the endoderm, while the ectoderm population has already moved in an alternative direction (Figure S1B). These high temporal resolution gene expression signatures suggest that expression programs associated with the three unique cell populations, representing early stages of each germ layer are established within a similar timeframe of hESC differentiation.

Global Transcriptional Dynamics Between hESCs and hESC-derived Cell Types

Based on these results, we selected day five as the optimal time point to capture early regulatory events in well-differentiated populations representing all three germ layers. To reduce heterogeneity, we used FACS to enrich populations based on previously reported surface markers (Figure S1C); populations isolated by FACS are referred to as dEC for the ectoderm, dME for the mesoderm and dEN for the endoderm. Expression analysis of the sorted populations confirms further enrichment for the desired populations (Figure 1E, Table S1).

We next expanded on our selected gene signature profiles by performing strand specific RNA-Seq on poly-A fractions from each day 5 differentiated FACS-isolated populations and undifferentiated HUES64 (Table S2). Hierarchical clustering based on the global expression profiles of each cell type reveals that the dME population is the most distantly related cell type and that dEN and dEC are more similar to each other than to dME or hESCs (Figure 1F). This was unexpected given that the dME and dEN populations are putatively derived through a common mesendoderm precursor stage (Figure 1B–C, Table S1) while the dEC does not upregulate markers associated with this stage (*EOMES, T, GSC;* Table S1). Overall, 14,196 RefSeq defined coding and non-coding transcripts (≈38% of defined transcripts) are expressed (FPKM >1) in at least one of the populations, with 11,579 (81.6% of the total number of transcripts detected within our cell types) being expressed in all three populations. Examining the overlap of genes expressed (FPKM >1) in each population reveals that the dME population exhibits expression of the largest number of unique genes (n=448, Figure 1G), such as *RUNX1* (FPKM: 3.4) and *HAND2* (FPKM: 17.8). Examining genes unique to pairs of the differentiated cell types also reveals that dEC and dME have the least in common (n=37, Figure 1G), while the dEC and dEN have the most number of transcripts in common (n=171, Figure 1G) consistent with our clustering analysis. Genes such as *PAX6* (dEC FPKM: 25.9, dEN FPKM: 5.6) and *NKX6.1* (dEC FPKM: 2.3, dEN FPKM: 3.3), which are each required for both brain (Ericson *et al.*, 1997) and pancreas development (Sander *et al.*, 1997), are expressed in both the dEC and dEN. Canonical markers of embryonic development, such as *FOXA2* (FPKM: 12.7) in the dEN and *EN1*

(FPKM: 5.8) in the dEC are restricted to their expected germ layers at our early stages (Table S2).

Notably, we also identified 1,296 splicing events (FDR=5%) as well as alternative promoter usage within our populations (Table S3) (Trapnell *et al.*, 2013). For example, we detected expression of multiple isoforms of *DNMT3B* ($p=5\times10^{-5}$). Expression of *DNMT3B* isoform 1 (NM_006892) was restricted to the undifferentiated hESCs (FPKM: 214.3), while the differentiated cell types predominantly express an alternative isoform, *DNMT3B* isoform 3 (NM_175849) (dEC FPKM: 33.9, dME FPKM: 14.2, dEN FPKM: 20.0) (Figure 1H). The presence of this isoform, as well as others, has previously been reported in more advanced stages of embryonic development as well as normal adult (Robertson *et al.*, 1999) and cancerous tissues (Ostler *et al.*, 2007). Our results suggest that this switch coincides with the exit from the pluripotent state, regardless of the specified lineage. We also identified expression of three *PITX2* isoforms, with differential splicing leading to different isoform expression between the dEN and dME (Table S3). In the chick *PITX2* is essential for heart looping and each isoform is responsible for executing distinct functions (Yu *et al.*, 2001). Taken together, this suggests that both transcript levels and isoform expression contribute to cellular identity.

Generation of Comprehensive Reference Epigenome Maps

To gain a more complete picture of the underlying molecular dynamics and investigate the regulatory events during the specification of the three germ layers, we collected approximately 12 million cells of the respective dEC, dME and dEN populations as well as HUES64. All samples were subjected to ChIP-Seq (H3K4me1, H3K4me3, H3K27me3, H3K27ac, H3K36me3, and H3K9me3) and whole genome bisulfite sequencing (WGBS) (Figure 2A), producing a total of 32 data sets with over 12 billion aligned reads (data are publicly available through the NIH Roadmap Epigenomics Project data repositories: http://www.roadmapepigenomics.org/, Table S4).

Integrative Analysis of Epigenetic State Transitions

After completing our basic quality control (see Extended Experimental Procedures and Table S4), we focused our analysis on previously identified informative chromatin states associated with various types of regulatory elements (Ernst *et al.*, 2011; Rada-Iglesias *et al.*, 2011), including the following specific combinations: H3K4me3+H3K27me3 (bivalent/poised promoter); H3K4me3+H3K27ac (active promoter); H3K4me3 (initiating promoter); H3K27me3+H3K4me1 (poised developmental enhancer); H3K4me1 (poised enhancer), H3K27ac+H3K4me1 (active enhancer); H3K27me3 (Polycomb-repressed) and H3K9me3 (heterochromatin). In addition, we segmented the WGBS data into three DNAme states: highly methylated regions (HMRs: >60%), intermediately methylated regions (IMRs: 11–60%) and unmethylated regions (UMRs: 0–10%). The latter differs from the highly methylated background of the genome and likely indicates functional importance as previously suggested (Stadler *et al.*, 2011). We next assigned each genomic region to one of the resulting states (Figure S2A, see Extended Experimental Procedures for details).

Across the four cell populations, we identified 268,062 genomic regions spanning 400 base pairs (bp) to 10 kilobases (kb) (Figure 2B) and covering a total of 450,058,678 bp (18.6% of the human genome) (Figure 2C), that were enriched for at least one of the 8 chromatin states and/or classified as an IMR or UMR in at least one of the cell types. Of the identified epigenetic states, H3K27me3-enriched regions and HMRs covered the most base pairs (Figure 2C) and the combination of H3K4me3 and H3K27me3 exhibits the highest CpG content (Figure 2D), as expected given the close association between H3K27me3 and high CpG density (Ku *et al.*, 2008). The majority of epigenetically dynamic regions are not located near promoters (6.8% +2kb to −500bp of the TSS- Promoters; 48.8% >50 kb upstream of TSS- Intergenic; 15.1% >500 bp downstream of TSS- Intragenic/Gene body) (Figure S2B). As expected, regions of open chromatin exhibit the highest median expression value (Figure 2E). However, overall we find that the majority (62–67%) of all epigenetic remodeling events are not directly linked to transcriptional changes based on the expression of the nearest gene.

The loss of H3K4 methylation (me1 and me3) is commonly associated with a transition to high DNAme (Figure 2F), which is most prominent in the dEN population and preferentially eliminated from genes involved in neural development (i.e. neural tube development $q=9.6\times10^{-12}$). We identified 4,639 proximal bivalent domains in hESCs and observe that 3,951 (85.1%) of these domains resolve their bivalent state in at least one hESC-derived cell type (Figure 2F, Figure S2C). When we specifically investigated the promoters of TF-encoding genes, we found that 463 of these promoters are in a bivalent state in hESCs, and 400 of them change in at least one differentiated cell type (Figure S2D). The majority transitions to H3K4me3-only or H3K27me3-only in a lineage-specific manner, as shown for *ISL1* (Figure S2E). In dME, H3K4me3 is gained at the *ISL1* locus while H3K27me3 is lost, leading to expression (FPKM: 14.3). The lineage specific dynamics in this region are interesting given that this gene has known roles in all three germ layers, although at later time points (Pfaff *et al.*, 1996; Ahlgren *et al.*, 1997; Cai *et al.*, 2003). Notably, in contrast to the limited overall association between many epigenomic dynamics and changes in expression, we found that a large proportion of these bivalent TFs (275) change their expression level during the differentiation (Figure S2D).

Pluripotent TF Binding is Linked to Chromatin Dynamics During Differentiation

To further explore potential regulators of chromatin dynamics during the exit from pluripotency, we performed ChIP-Seq for OCT4, NANOG and SOX2 in HUES64 (Figure S2F and G). We found that regions bound by all three factors (n=1,556), by SOX2-only (n=923) or by NANOG-only (n=14,531) are frequently associated with inter- and intragenic regions (Figure S2H–S2J, **top**). In contrast, regions bound by OCT4-only (n=8,599) are more frequently associated with promoter regions (Figure S2H). Examination of regions bound by OCT4, NANOG and SOX2 in hESCs showed H3K4me1 regions enriched for OCT4 binding sites frequently become HMRs in all three differentiated cell types whereas NANOG and SOX2 sites are more prone to change to an HMR state in dME (Figure 2I). In general, many regions associated with open chromatin that are bound by NANOG are more likely to retain this state in dEN compared to dME and dEC (Figure 2I). We also found that regions enriched for H3K27ac in hESCs that maintain this state in dEN or dEC are likely to

be bound by SOX2 and NANOG. This is in agreement with the reported role of SOX2 during ectoderm development and differentiation (Wang *et al.*, 2012), but also supports our observation that SOX2 expression is maintained in the dEN. Motif enrichment analysis detected the GATA3 motif in regions bound by OCT4 and SOX2 that transition to an active state in dEC. Furthermore, we found that regions bound by OCT4, NANOG and SOX2 that gain an active mark in dEC are enriched for the motifs PAX9, p63 and STATs (Table S5). Examining epigenetic dynamics at sites of OCT4, NANOG and SOX2 binding further supports the observation that some pluripotency associated TFs are also involved in the downstream specification.

Gain of DNAme Occurs at Open Chromatin enriched for TF motifs

We next utilized the WGBS data that cover approximately 26 Million CpGs (at ≥5 coverage) across all four cell types. Hierarchical clustering analysis of the WGBS data, which included human adult liver and hippocampus for comparison, revealed that the pluripotent hESCs and the hESC-derived cell types form a separate cluster arm with respect to the somatic tissues (Figure 3A).

We determined DMRs defined as exhibiting a significant ($p \leq 0.05$) minimal difference of CpG methylation level of 0.1 among our four cell types. The majority of all DMRs occur at CpG-poor intergenic regions in line with previous reports (Figure 3B, **bottom**) (Stadler et al., 2011). The dEN exhibits more than twice the number of regions that gain DNAme compared to dEC and dME (Figure 3B, **top**). Interestingly, only 65 of the total number of DMRs identified are shared between all three populations. However, reaffirming that our populations are depleted of pluripotent cells, this group of DMRs includes the regulatory region of OCT4 (Figure S3A). In line with the small number of shared regions, more than 60% of regions that gain DNAme are lineage specific (Figure 3C) and include loci such as *SMAD3* (dEC), *CTNNA3* (dME) and *FOXA2* (dEN). *FOXA2* has an upstream CGI that exhibits gain of DNAme (Figure 3D), and transcription in dEN is initiated downstream of this DMR at an alternative TSS, suggesting that TSS usage may be regulated, stabilized or reflected by DNAme (Maunakea *et al.*, 2010).

We find significant enrichment of various TF motifs as DNAme targets upon differentiation (Table S6), which has some analogy to the gain of methylation observed at myeloid targets in the lymphoid lineage *in vivo* (Ji *et al.*, 2010; Deaton *et al.*, 2011; Bock *et al.*, 2012). To extend this observation, we examined the DNAme state at regions bound by SOX2, OCT4 and NANOG in hESCs. For example, two regions 20 kb downstream of *DBX1*, a gene associated with early neural specification, are bound by all three TFs and gain DNAme in dME and dEN. In contrast, this region maintains low levels of DNAme in dEC, which has activated transcription of *DBX1* (Figure 3E). We generally find that co-bound sites gain DNAme in the dME and dEN, but not dEC (Figure S3B). Further supporting the functional relevance of these dynamics, we find that regions with gain DNAme frequently coincide with DNAseI hypersensitive sites (Figure S3C) (Thurman *et al.*, 2012). While transcriptional silencing was infrequently correlated with gain of DNAme at distal elements (Figure 3F, **left**), the promoters that gain DNAme in dEC and dME, are associated with a decrease in expression as expected (Figure 3F, **right**).

In examining the chromatin state of regions that gain DNAme during differentiation, we find that most regions exhibited enrichment of one or more histone modifications in hESCs (Figure 3G). These results confirm that in particular distal regulatory elements show highly dynamic regulation of DNA methylation during specification.

Loss of DNA Methylation is Biased Towards dEC

Loss of DNAme is asymmetric between the three populations (Figure 4A, **top**) and occurs in a more lineage specific fashion than gain (Figure 4B). However, loss also occurs mainly at intergenic regions (Figures 3B, **bottom** and 4A, **bottom**). Notably, the dEC has the most DMRs and many were associated with neuronal gene categories (for instance: neural tube development, $q=3.13\times10^{-13}$). This includes the ectodermal TF *POU3F1*, which has a bivalent promoter in hESCs, resolves to a H3K4me3-only state (Figure S4A) and exhibits transcriptional activation in dEC (Table S2). Chromatin remodeling and activation at this locus coincides with specific loss of DNAme at a putative regulatory element downstream of the 3'UTR of this gene in dEC (Figure 4C). On a global scale, an immediate correspondence between loss of DNAme and expression, such as that observed at *POU3F1*, occurs at about half the regions (Figure 4D). More than 70% of DMRs that lose DNAme during differentiation are enriched for one of our profiled histone modifications in particular H3K4me1 or H3K27ac (Figure 4E).

Taken together, our hESC differentiation system reveals several interesting DNAme dynamics, including the lineage specific silencing of regulatory regions in default or alternative lineages. The asymmetric loss may also explain why our chromatin state analysis revealed fewer regions that gained H3K4me3 in the dEC population.

Gain of H3K27ac Reveals Putative Regulatory Elements

In addition to methylation on H3K4, open chromatin is also demarcated by enrichment of H3K27ac. It has also been suggested that the combination of H3K4me1 and H3K27ac at distal regions identifies active enhancer elements, while H3K4me1 and H3K27me3 corresponds to poised enhancer elements (Rada-Iglesias *et al.*, 2011). To extend these observations, we focused specifically on regions that gain H3K27ac during differentiation and found that more than half of the identified regions are HMRs in hESCs (Figure 5A), while another large fraction is enriched for H3K4me1 in hESCs (Figure 5A). The majority of regions that gain H3K27ac are intergenic (Figure S5A), as shown for the *RUNX1* locus (Figure 5A, B).

We next placed each region into one of three distinct categories (repressed, poised, open) based on their state in hESCs, and subsequently performed gene set enrichment analysis using the GREAT toolbox (Figure 5C) (McLean *et al.*, 2010). This analysis reveals enhancer dynamics in line with the lineage specific differentiation trajectory for dEC and dME (Figure 5C). In contrast, the dEN population shows an unexpected enrichment for early neuronal genes (e.g. neural tube development, Figure 5C). This observation is consistent with the correlation that we reported between our dEC and dEN RNA-Seq data, suggesting that similar networks are induced in the early stages of both our ectoderm and endoderm specification (van Arensbergen *et al.*, 2010).

Moreover, we find strong enrichment of downstream effector genes of the TGFβ, VEGF and BMP pathways in dME, directly reflecting the signaling cascades that were stimulated to induce the respective differentiation. In dEN we find enrichment of genes involved in WNT/β- CATENIN and retinoic acid (RA) signaling (Figure 5C). While we did not use RA, this signaling cascade has previously been implicated in endodermal tissue development including pharyngeal and pancreatic cell types (Wendling *et al.*, 2000; Ostrom *et al.*, 2008). Concordantly, we also find high levels of SMAD3 motif enrichment in the repressed dME and dEN, particularly in the poised putative enhancer populations (Figure 5D). Similarly, we observe enrichment of key lineage specific TF motifs such as the ZIC family proteins in dEC, TBX5 in dME and SRF in dEN. Interestingly, we also find the FOXA2 motif highly overrepresented in dEN where the factor is active, and also dEC where the factor is inactive but becomes expressed at a later stage of neural differentiation (Kriks *et al.*, 2011), but not in dME (Figure 5D).

Acquisition of H3K4me1 without Transcriptional Activation Suggests Epigenetic Priming

Many regions that exhibit high DNAme in hESCs and transition to H3K4me1 in one lineage remain HMRs in the two alternative cell types (Figure 2F, Figure 6A). Similar to the regions showing dynamic DNAme during differentiation, these regions are typically intergenic and >10kb from the nearest TSS (Figure 6B). GREAT analysis of these regions shows a strong enrichment for categories associated with brain development such as cerebellum morphogenesis in dEC ($q<10^{-30}$), TGFβ pathway targets ($q<10^{-10}$) in dME and suppression of EMT in dEN ($q<0.0001$). To understand if regions that gain H3K4me1 in our system are associated with somatic identity, we took advantage of published microarray data for 24 human tissues and determined genes upregulated in these tissues with respect to hESCs (termed, Tissue Atlas, see Extended Experimental Procedures). Reaffirming the relevance of our dynamics, we found regions that gain H3K4me1 in dEC are associated with fetal brain and specific cell types found within the adult brain (Figure 6C) based on region association with its nearest gene. The dME H3K4me1 pattern was associated with a range of interrogated tissues, such as heart, spinal cord and stomach, which may be due to heterogeneity of the tissues collected (Figure 6C). The dEN associations were interesting given that, as with the RNA-Seq and H3K27ac trends, H3K4me1 was again associated with brain-related categories (Figure 6C).

Overall, less than half of the genes that gain H3K4me1 exhibit immediate transcriptional changes (Figure 6D). *CYP2A6* and *CYP2A7* (Figure 6E) are representative examples that do not show a corresponding change in expression, while *LMO2* does (Figure 6F). To investigate these regions in more detail, we carried out motif enrichment analysis and found lineage specific enrichment of TF motifs near regions that gain H3K4me1. While the FOXA2 motif is enriched in all three cell types, the DBX1 motif is associated with the gain of H3K4me1 in dEC (Figure 6G), which coincides with its transcriptional activation in this cell type (FPKM: 5.36). Conversely, the GLI3, HIC1 and CTF1 motifs are strongly enriched at regions that gain H3K4me1 in dEN (Figure 6G).

To further assess if this DNAme to H3K4me1 switch acts as a priming event, we differentiated the HUES64 endoderm population for five additional days in the presence of

BMP4 and FGF2, leading to HNF4α positive hepatoblast-like (dHep) cells (Table S2, Figure S6A). Interestingly, of the motifs enriched in dEN that gain H3K4me1, HIC1, KLF4 and CTF1 (Figure 6G), several of these genes become expressed at the next stage of differentiation (Figure 6H). Lastly, 1,346 of these putatively primed regions are enriched for the active enhancer mark H3K27ac in human liver (Figure 6I).

Loss of DNA methylation and acquisition of H3K27me3 at putative regulatory elements

More surprisingly, we observe intergenic regions that switch from high DNAme to H3K27me3 (n=3,985 in dEN) (Figure 7A). This transition frequently occurred within CpG poor, distal regions, which is distinct from the common CpG island-centric targets of Polycomb Repressive Complex 2 and H3K27me3 (Figure 7B). This switch is highly lineage specific and DNAme is generally retained in the alternative two cell types (Figure 7C, Figure 2F).

Motif enrichment analysis, combined with the evaluation of publicly available TF binding site (TFBS) data from the ENCODE project, indicated that many regions exhibiting this transition in dEN were near binding sites of the pioneering factor FOXA2. This TF has putative roles in chromatin decompaction, but its distinct functions and limitations remain somewhat unclear (Li *et al.*, 2012). To investigate this association, we performed ChIP-Seq for FOXA2 in the endoderm population. This analysis reveals that FOXA2 binding sites frequently overlap with regions that transition from HMR to H3K27me3 (Figure 7D). We also confirmed that gain of H3K27me3 at dEN FOXA2 binding sites occurs predominantly in dEN, and not dEC or dME (Figure 7E). A notable example of this transition can be seen at the *ALB* locus, where H3K27me3 is gained at *AFP* and *AFM*, proximal to FOXA2 binding sites (Figure 7F). This mark is not found in primary liver tissue, suggesting it represents a transient state (Figure 7F). Many regions that exhibit this transition are required for later stages of development as with *AFP* and *AFM* or *HBB1* in the dME. Importantly, the majority of these regions do not yet exhibit significant increases in expression (Figure S7A).

A previous report found that FOXA1/FOXA2 could bind to regions exhibiting DNAme (Serandour *et al.*, 2011), which is not a characteristic shared by all TFs (Gifford and Meissner, 2012). Regions bound by these factors subsequently lost DNAme and gained euchromatic histone modifications in our populations. We therefore compared DNAme at FOXA2 binding sites in hESCs to dEN and found a slight reduction in specifically in the dEN (Figure 7G). To more directly assess this relationship, we interrogated the DNAme state of regions isolated by FOXA2-ChIP-Bisulfite sequencing in dEN (Brinkman *et al.*, 2012). Interestingly, we saw a major depletion of DNAme at sites isolated by FOXA2-ChIP (Figure 7G). To determine if these regions exhibit transcriptional activation after further differentiation we examined again our dHep RNA-Seq data and found that 50 genes, which were bound by FOXA2 and gained H3K27me3 in dEN, increased their expression (Figure 7H, Table S2). We also find H3K27ac enrichment at 197 loci in the human liver that had experienced the gain of H3K27me3 in dEN (Figure 7I).

Discussion

Using directed differentiation of hESCs to three distinct, FACS-enriched populations, representing early stages of embryonic development, we provide an extensive set of new data and many insights on the transcriptional and epigenetic dynamics that occur during human *in vitro* lineage specification.

Among other things we describe two very interesting, but distinct lineage specific dynamics from high DNAme to H3K4me1 or H3K27me3. These transitions occur at many sites that do not significantly change gene expression during our early stages of differentiation. Notably, we made similar observations for H3K4 methylation during the early stages of reprogramming to an iPSC state (Koche *et al.*, 2011), suggesting that this type of epigenetic priming event might be common. At this point however it is not clear whether these events reflect a regulatory mechanism to facilitate timely activation upon differentiation or indicate the absence of a critical co-factor necessary for complete transcriptional activation. We also cannot rule out that a subset of the observed priming events are due to heterogeneity in the cell population that are not detected by our RNA-Seq. Our observation that high DNAme switches to H3K27me3 enrichment in distal, CpG-poor regions is even more interesting. It remains to be tested whether targeted loss of DNAme at these regions causes a default gain of H3K27me3 in the absence of additional co-factors due to underlying sequence context (Mendenhall *et al.*, 2010) or represents a more active recruitment event and regulatory mechanism. It is also possible that H3K27me3 gain at distal regions is due to genomic conformation changes and reflects H3K27me3 spreading in three dimensions. It was recently reported that the combination of H3K27me3 enrichment and a nearby nucleosome-depleted region creates sites amenable to TF binding (Taberlay *et al.*, 2011). Based on these results, one may speculate that specific TFs, such as FOXA2, exert chromatin decompaction functions resulting in loss of DNAme leading to gain of H3K27me3, which creates a platform for subsequent binding of other TFs that cannot directly remodel a heterochromatic state, but instead function in transcription machinery assembly and transcriptional activation.

In conclusion, our data provide new insights on transcriptional and epigenetic dynamics during hESC specification and represent valuable reference maps for many applications, including regenerative biology and the study of human developmental biology.

Experimental Procedures

For full details, see Extended Experimental Procedures.

Cell Culture

Human ESCs were cultured as previously described (Bock *et al.*, 2011). All details regarding the differentiation conditions are in the Extended Experimental Procedures.

Gene expression signatures

RNA was hybridized to a custom probe set, processed using the Nanostring prep station, imaged using the Nanostring nCounter and analyzed as previously described (Bock *et al.*, 2011).

Sequencing Library Construction

For WGBS sequencing libraries, Genomic DNA was fragmented using a Covaris S2 sonicator. DNA fragments were cleaned-up, end-repaired, A-tailed, and ligated with methylated paired-end adapters (purchased from ATDBio).

For ChIP-Seq libraries, ChIP-isolated DNA was end-repaired, A-tailed and ligated to barcoded Illumina adaptors, and the library was amplified using PFU Ultra II Hotstart Master Mix (Agilent)

For RNA-Seq libraries, polyadenylated RNA was isolated using Oligo dT beads (Invitrogen) and fragmented to 200–600 base pairs, and then ligated to RNA adaptors using T4 RNA Ligase, (NEB), preserving strand of origin information.

ChIP-BS libraries were constructed as previously described (Brinkman *et al.*, 2012).

Sequencing Data Processing and Analyses

Details regarding the WGBS, ChIP-Seq, ChIP-bisulfite Seq and RNA-Seq bioinformatics analysis can be found in the Extended Experimental Procedures. All data was aligned to hg19, and the accession numbers for each data set are included in Table S4.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Kendell Clement for support of the WGBS data visualization, Loyal Goff for RNA-Seq visualization, as well as Zachary Smith and Jing Liao, for critical reading of the manuscript, and the SCRB FACS Core for advice regarding FACS analysis. We also thank other members of the Meissner Lab and Epigenomics Platform at the Broad Institute for helpful discussion. A.M. is supported by the Pew Charitable Trusts and is a New York Stem Cell Foundation (NYSCF) Robertson Investigator. The work was funded by the US National Institutes of Health (NIH) grants (U01ES017155 and P01GM099117) and the NYSCF.

References

Ahlgren U, Pfaff SL, Jessell TM, Edlund T, Edlund H. Independent requirement for ISL1 in formation of pancreatic mesenchyme and islet cells. Nature. 1997; 385:257–260. [PubMed: 9000074]

Bock C, Beerman I, Lien WH, Smith ZD, Gu H, Boyle P, Gnirke A, Fuchs E, Rossi DJ, Meissner A. DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. Mol Cell. 2012; 47:633–647. [PubMed: 22841485]

Bock C, Kiskinis E, Verstappen G, Gu H, Boulting G, Smith ZD, Ziller M, Croft GF, Amoroso MW, Oakley DH, et al. Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. Cell. 2011; 144:439–452. [PubMed: 21295703]

Brinkman AB, Gu H, Bartels SJ, Zhang Y, Matarese F, Simmer F, Marks H, Bock C, Gnirke A, Meissner A, et al. Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation

of chromatin and DNA methylation cross-talk. Genome Res. 2012; 22:1128–1138. [PubMed: 22466170]

Cai CL, Liang X, Shi Y, Chu PH, Pfaff SL, Chen J, Evans S. Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. Dev Cell. 2003; 5:877–889. [PubMed: 14667410]

Chambers SM, Qi Y, Mica Y, Lee G, Zhang XJ, Niu L, Bilsland J, Cao L, Stevens E, Whiting P, et al. Combined small-molecule inhibition accelerates developmental timing and converts human pluripotent stem cells into nociceptors. Nat Biotechnol. 2012; 30:715–720. [PubMed: 22750882]

Chen YF, Tseng CY, Wang HW, Kuo HC, Yang VW, Lee OK. Rapid generation of mature hepatocyte-like cells from human induced pluripotent stem cells by an efficient three-step protocol. Hepatology. 2012; 55:1193–1203. [PubMed: 22095466]

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 2010; 107:21931–21936. [PubMed: 21106759]

Deaton AM, Webb S, Kerr AR, Illingworth RS, Guy J, Andrews R, Bird A. Cell type-specific DNA methylation at intragenic CpG islands in the immune system. Genome Res. 2011; 21:1074–1086. [PubMed: 21628449]

DeLaForest A, Nagaoka M, Si-Tayeb K, Noto FK, Konopka G, Battle MA, Duncan SA. HNF4A is essential for specification of hepatic progenitors from human pluripotent stem cells. Development. 2011; 138:4143–4153. [PubMed: 21852396]

Ericson J, Rashbass P, Schedl A, Brenner-Morton S, Kawakami A, van Heyningen V, Jessell TM, Briscoe J. Pax6 controls progenitor cell identity and neuronal fate in response to graded Shh signaling. Cell. 1997; 90:169–180. [PubMed: 9230312]

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011; 473:43–49. [PubMed: 21441907]

Evseenko D, Zhu Y, Schenke-Layland K, Kuo J, Latour B, Ge S, Scholes J, Dravid G, Li X, MacLellan WR, et al. Mapping the first stages of mesoderm commitment during differentiation of human embryonic stem cells. Proc Natl Acad Sci U S A. 2010; 107:13742–13747. [PubMed: 20643952]

Gifford CA, Meissner A. Epigenetic obstacles encountered by transcription factors: reprogramming against all odds. Curr Opin Genet Dev. 2012; 22:409–415. [PubMed: 22922161]

Hay DC, Fletcher J, Payne C, Terrace JD, Gallagher RC, Snoeys J, Black JR, Wojtacha D, Samuel K, Hannoun Z, et al. Highly efficient differentiation of hESCs to functional hepatic endoderm requires ActivinA and Wnt3a signaling. Proc Natl Acad Sci U S A. 2008; 105:12301–12306. [PubMed: 18719101]

Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature. 2009; 459:108–112. [PubMed: 19295514]

Ji H, Ehrlich LI, Seita J, Murakami P, Doi A, Lindau P, Lee H, Aryee MJ, Irizarry RA, Kim K, et al. Comprehensive methylome map of lineage commitment from haematopoietic progenitors. Nature. 2010; 467:338–342. [PubMed: 20720541]

Koche RP, Smith ZD, Adli M, Gu H, Ku M, Gnirke A, Bernstein BE, Meissner A. Reprogramming factor expression initiates widespread targeted chromatin remodeling. Cell Stem Cell. 2011; 8:96–105. [PubMed: 21211784]

Kriks S, Shim JW, Piao J, Ganat YM, Wakeman DR, Xie Z, Carrillo-Reid L, Auyeung G, Antonacci C, Buch A, et al. Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson's disease. Nature. 2011; 480:547–551. [PubMed: 22056989]

Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum C, Xie X, Chi AS, et al. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet. 2008; 4:e1000242. [PubMed: 18974828]

Lee G, Chambers SM, Tomishima MJ, Studer L. Derivation of neural crest cells from human pluripotent stem cells. Nat Protoc. 2010; 5:688–701. [PubMed: 20360764]

Li E. Chromatin modification and epigenetic reprogramming in mammalian development. Nat Rev Genet. 2002; 3:662–673. [PubMed: 12209141]

Li Z, Gadue P, Chen K, Jiao Y, Tuteja G, Schug J, Li W, Kaestner KH. Foxa2 and H2A.Z mediate nucleosome depletion during embryonic stem cell differentiation. Cell. 2012; 151:1608–1616. [PubMed: 23260146]

Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature. 2010; 466:253–257. [PubMed: 20613842]

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010; 28:495–501. [PubMed: 20436461]

Mendenhall EM, Koche RP, Truong T, Zhou VW, Issac B, Chi AS, Ku M, Bernstein BE. GC-rich sequence elements recruit PRC2 in mammalian ES cells. PLoS Genet. 2010; 6:e1001244. [PubMed: 21170310]

Ostler KR, Davis EM, Payne SL, Gosalia BB, Exposito-Cespedes J, Le Beau MM, Godley LA. Cancer cells express aberrant DNMT3B transcripts encoding truncated proteins. Oncogene. 2007; 26:5553–5563. [PubMed: 17353906]

Ostrom M, Loffler KA, Edfalk S, Selander L, Dahl U, Ricordi C, Jeon J, Correa-Medina M, Diez J, Edlund H. Retinoic acid promotes the generation of pancreatic endocrine progenitor cells and their further differentiation into beta-cells. PLoS One. 2008; 3:e2841. [PubMed: 18665267]

Pfaff SL, Mendelsohn M, Stewart CL, Edlund T, Jessell TM. Requirement for LIM homeobox gene Isl1 in motor neuron generation reveals a motor neuron-dependent step in interneuron differentiation. Cell. 1996; 84:309–320. [PubMed: 8565076]

Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. A unique chromatin signature uncovers early developmental enhancers in humans. Nature. 2011; 470:279–283. [PubMed: 21160473]

Robertson KD, Uzvolgyi E, Liang G, Talmadge C, Sumegi J, Gonzales FA, Jones PA. The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors. Nucleic Acids Res. 1999; 27:2291–2298. [PubMed: 10325416]

Sander M, Neubuser A, Kalamaras J, Ee HC, Martin GR, German MS. Genetic analysis reveals that PAX6 is required for normal transcription of pancreatic hormone genes and islet development. Genes Dev. 1997; 11:1662–1673. [PubMed: 9224716]

Serandour AA, Avner S, Percevault F, Demay F, Bizot M, Lucchetti-Miganeh C, Barloy-Hubler F, Brown M, Lupien M, Metivier R, et al. Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. Genome Res. 2011; 21:555–565. [PubMed: 21233399]

Smith ZD, Meissner A. DNA methylation: roles in mammalian development. Nat Rev Genet. 2013; 14:204–220. [PubMed: 23400093]

Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D, et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature. 2011; 480:490–495. [PubMed: 22170606]

Taberlay PC, Kelly TK, Liu CC, You JS, De Carvalho DD, Miranda TB, Zhou XJ, Liang G, Jones PA. Polycomb-repressed genes have permissive enhancers that initiate reprogramming. Cell. 2011; 147:1283–1294. [PubMed: 22153073]

Teo AK, Arnold SJ, Trotter MW, Brown S, Ang LT, Chng Z, Robertson EJ, Dunn NR, Vallier L. Pluripotency factors regulate definitive endoderm specification through eomesodermin. Genes Dev. 2011; 25:238–250. [PubMed: 21245162]

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, et al. The accessible chromatin landscape of the human genome. Nature. 2012; 489:75–82. [PubMed: 22955617]

Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol. 2013; 31:46–53. [PubMed: 23222703]

van Arensbergen J, Garcia-Hurtado J, Moran I, Maestro MA, Xu X, Van de Casteele M, Skoudy AL, Palassini M, Heimberg H, Ferrer J. Derepression of Polycomb targets during pancreatic

organogenesis allows insulin-producing beta-cells to adopt a neural gene activity program. Genome Res. 2010; 20:722–732. [PubMed: 20395405]

Wamstad JA, Alexander JM, Truty RM, Shrikumar A, Li F, Eilertson KE, Ding H, Wylie JN, Pico AR, Capra JA, et al. Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. Cell. 2012; 151:206–220. [PubMed: 22981692]

Wang Z, Oron E, Nelson B, Razis S, Ivanova N. Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. Cell Stem Cell. 2012; 10:440–454. [PubMed: 22482508]

Wei H, Tan G, Manasi, Qiu S, Kong G, Yong P, Koh C, Ooi TH, Lim SY, Wong P, et al. One-step derivation of cardiomyocytes and mesenchymal stem cells from human pluripotent stem cells. Stem Cell Res. 2012; 9:87–100. [PubMed: 22683798]

Wendling O, Dennefeld C, Chambon P, Mark M. Retinoid signaling is essential for patterning the endoderm of the third and fourth pharyngeal arches. Development. 2000; 127:1553–1562. [PubMed: 10725232]

Yu X, St Amand TR, Wang S, Li G, Zhang Y, Hu YP, Nguyen L, Qiu MS, Chen YP. Differential expression and functional analysis of Pitx2 isoforms in regulation of heart looping in the chick. Development. 2001; 128:1005–1013. [PubMed: 11222154]

Zhang JA, Mortazavi A, Williams BA, Wold BJ, Rothenberg EV. Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. Cell. 2012; 149:467–482. [PubMed: 22500808]

Zhou VW, Goren A, Bernstein BE. Charting histone modifications and the functional organization of mammalian genomes. Nat Rev Genet. 2011; 12:7–18. [PubMed: 21116306]

Highlights

- Epigenetic and transcriptional dynamics in hESCs and hESC-derived populations
- Lineage specific remodeling at regions bound by OCT4, SOX2 and NANOG in hESCs
- Germ layer specific switch to H3K4me1 or H3K27me3 at sites of high DNA methylation
- Epigenetic dynamics frequently precede transcriptional activation

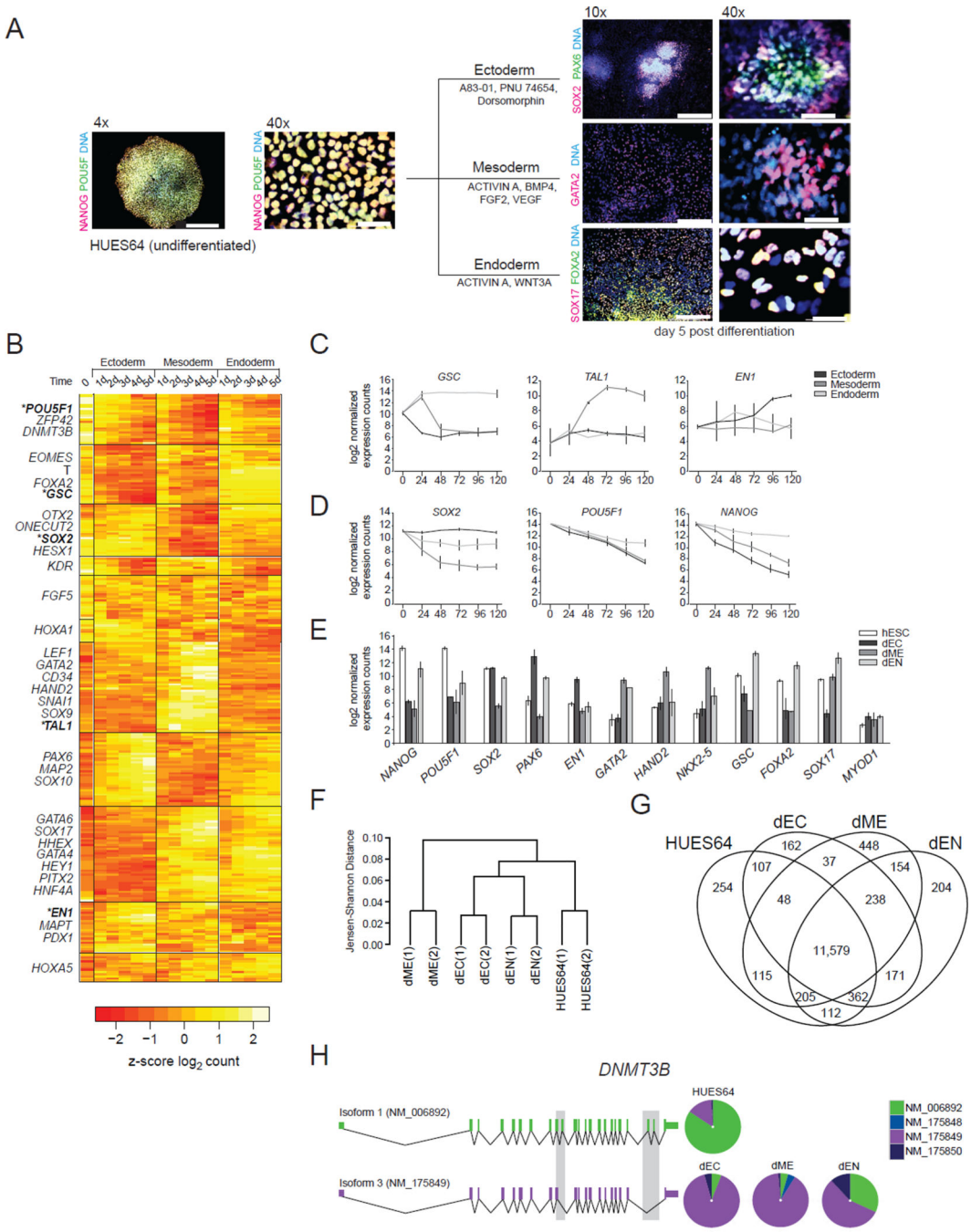

Figure 1. Generation and Characterization of hESCs and hESC-Derived Cell Types

A. Left Low (4×) and high (40×) magnification overlaid immunofluorescent images of the undifferentiated human embryonic stem cell (hESC) line HUES64 stained with OCT4 (POU5F1) and NANOG antibodies. **Right:** Established directed (2-dimensional) differentiation conditions were used to generate representative populations of the three embryonic germ layers: hESC-derived ectoderm, hESC-derived mesoderm and hESC-derived endoderm. Cells were fixed and stained after 5 days of differentiation with the indicated antibodies. Representative overlaid images at low (10×) and high (40×)

magnification are shown. DNA was stained with Hoechst 33342 in all images. Scale bars = 200μm (4×), 100 μm (10×) and 30 μm (40×).

B. NanoString nCounter expression data (z-score $\log_2$ expression value of two biological replicates) for a time-course of *in vitro* differentiation using the conditions shown in panel **A**. 541 genes were profiled and 268 changing by more than 0.7 are displayed. Selected lineage-specific genes are shown on the left for each category that was identified based on hierarchical clustering (see Table S1 for all). The average $\log_2$ expression value of two biological replicates is displayed. Error bars represent one standard deviation (SD).

C. The average $\log_2$ expression values of two biological replicates of lineage specific genes highlighted in panel B are shown. Error bars represent one SD. If no error is evident, SD < 0.5 $\log_2$ expression units.

D. The average $\log_2$ expression values of two biological replicates of pluripotent genes highlighted in panel B are shown. Error bars represent one SD. If no error is evident, SD < 0.5 $\log_2$ expression units.

E. NanoString nCounter profiling of FACS-isolated ectoderm (dEC), mesoderm (dME), and endoderm (dEN). Expression levels for *MYOD1* (**right**) are included as a negative control. The average $\log_2$ expression value of two biological replicates is shown. Error bars represent one SD. If no error is evident, SD < 0.5 $\log_2$ expression units.

F. Hierarchical clustering of global gene expression profiles as measured by strand-specific RNA-Seq for biological replicates of HUES64 and dEC, dME, and dEN is shown as a dendrogram. Pairwise distances between the replicates were measured using the Jensen-Shannon distance metric.

G. Venn diagram illustrating unique and overlapping genes with expression (FPKM > 1) in HUES64 and the FACS-isolated directed differentiation conditions are shown.

H. Differential splicing of *DNMT3B* in response to directed differentiation. Relative expression of isoforms 1 (NM_006892, green) and 3 (NM_175849, purple) as measured by RNA-Seq are shown on the right.

See also Figure S1

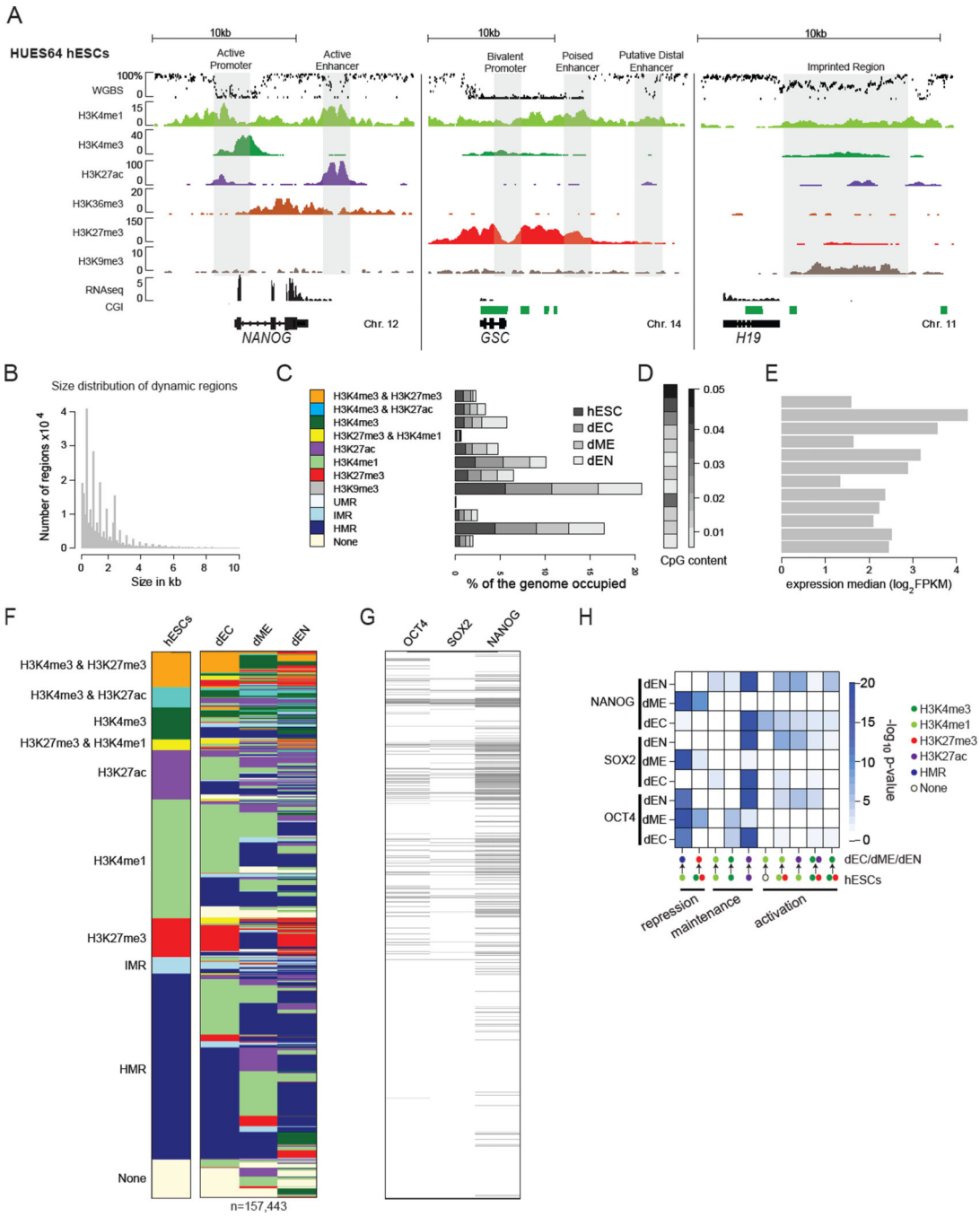

Figure 2. Epigenetic Remodeling Is Lineage Specific During Directed Differentiation

A. WGBS (% methylation), ChIP-Seq (read count normalized to 10 million reads) and RNA-Seq (FPKM, read count normalized) for the undifferentiated hESC line HUES64 at three loci: *NANOG* (chr12:7,935,038-7,957,818), *GSC* (chr14:95,230,449-95,250,241), and *H19* (chr11:2,015,282-2,027,359). CpG Islands (CGI) are indicated in green.

B. Size distribution of genomic regions enriched for at least one of our 6 histone modifications in at least one cell type (hESC, dEC, dME, dEN) and/or classified as UMR or IMR in at least one cell type (n=297,617).

C. Definition of epigenetic states used in this study and the genomic space occupied by these in the four cell types under study.

D. Median CpG content of the genomic regions in distinct epigenetic states defined in panel **C**.

E. Median expression level of epigenetic states used in this study (**C**) based on assignment of each region to the nearest RefSeq gene. Median was computed over the states in all four cell types and the corresponding expression profile.

F. Epigenetic state map of regions enriched for one of five histone modifications in at least one cell type or classified as UMR/IMR in at least one cell type and changing its epigenetic state upon differentiation in at least one cell type (n=157,433). State definitions are listed in panel **C**.

G. Regions bound by NANOG, OCT4, SOX2, as determined by ChIP-Seq and organized using the chromatin states in **2F**.

H. Enrichment of OCT4, SOX2 and NANOG within various classes of dynamic genomic regions changing upon differentiation of hESC, computed relative to all regions exhibiting the particular epigenetic state change in other cell types. Epigenetic dynamics are categorized into three major classes: repression (loss of H3K4me3 or H3K4me1 and acquisition of H3K27me3 or DNAme), maintenance of open chromatin marks (H3K4me3, H3K4me1, H3K27ac) and activation of previously repressed states.

See also Figure S2.

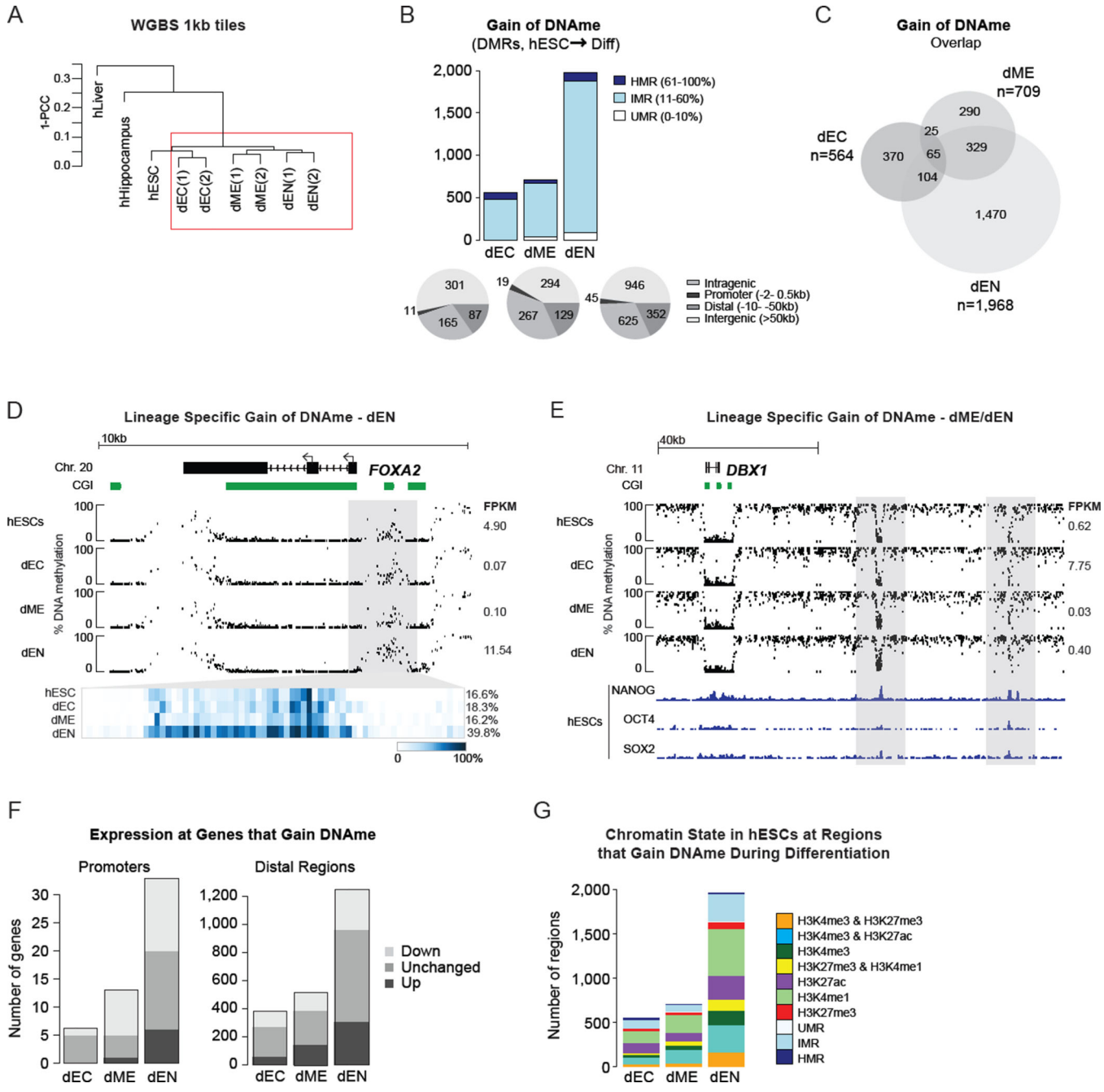

Figure 3. Global DNA Methylation Dynamics – Gain of DNAme A

A. Hierarchical clustering of hESCs, hESC-derived populations (dEC, dME and dEN), human adult hippocampus and human adult liver based on mean DNAme levels of 1kb tiles across the human genome using Pearson Correlation Coefficient (PCC). Y-axis indicates sample distance in terms of 1 minus PCC. Red box indicates cell types interrogated in this study.

B. Regions that significantly ($p \leq 0.05$) increase their DNAme levels by at least 0.1 between hESCs and the differentiated cell types. The color code indicates the DNAme state found in hESCs. Bottom: Genomic features associated with DMRs gaining DNAme in each of the

differentiated cell types based on RefSeq gene annotation and de novo discovered promoters by RNA-Seq.

C. The overlap of these differentially methylated regions (DMRs) that increase their DNAme level in the three hESC-derived populations.

D. DNAme levels and RNA-Seq expression values of *FOXA2* (chr20:22,559,343-22,571,189) in hESCs and differentiated cell types. The heat map below shows the DNAme values of individual CpGs within the highlighted region. The average DNAme value for the entire highlighted region is shown on the right in red. CpG islands (CGI) are shown as green bars. Expression values (FPKM) are displayed on the right. The arrows indicate two known TSSs.

E. DNAme levels and OCT4, SOX2 and NANOG ChIP-Seq at the *DBX1* locus (chr11:20,169,548-20,277,940).

F. Distal elements (left) and Promoters (right) that gain DNAme separated by the changes in FPKM at associated genes.

G. Chromatin state in hESCs at regions that gain DNAme during differentiation. Regions devoid of any detected chromatin marks are categorized according to their DNA methylation state in hESCs.

See also Figure S3.

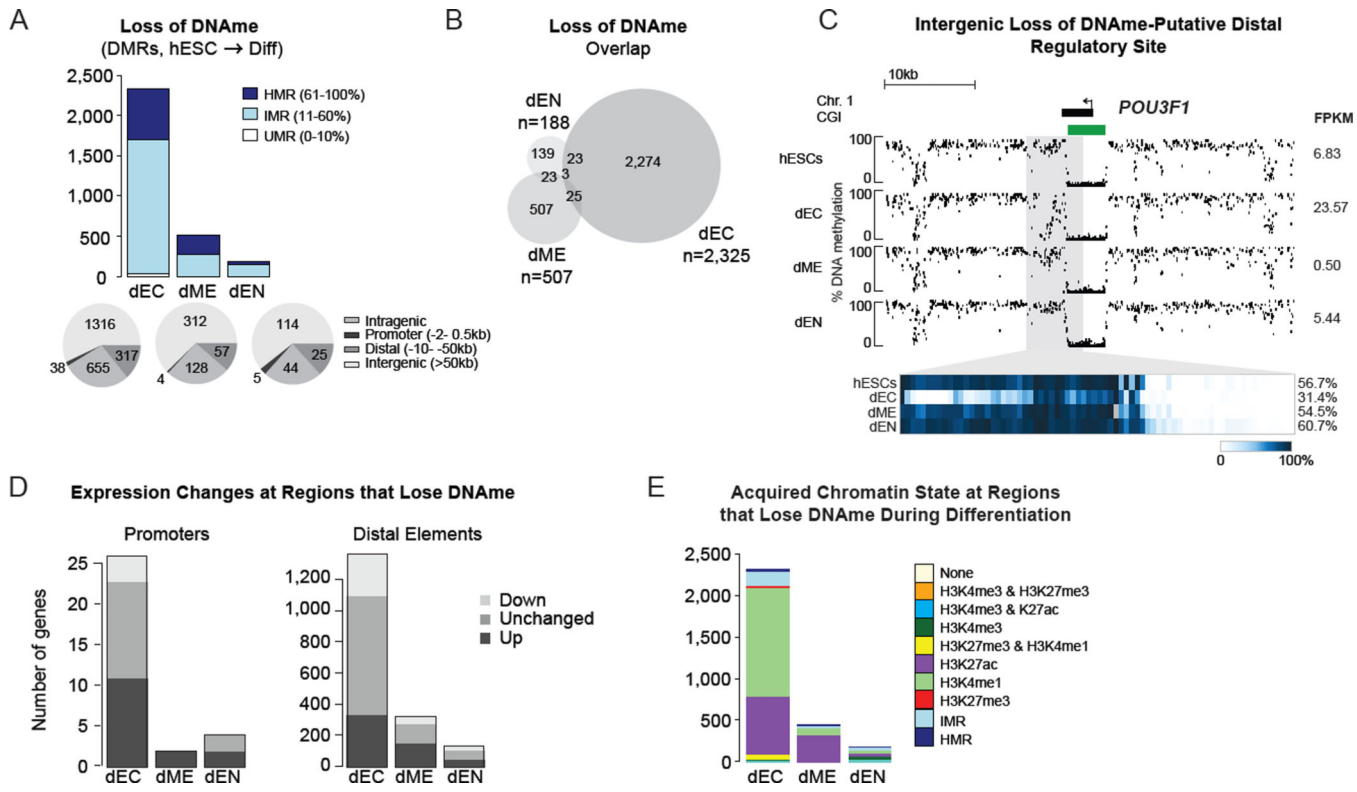

Figure 4. Global DNA Methylation Dynamics – Loss of DNAme

A. Regions that significantly ($p \leq 0.05$) decrease their DNAme levels by at least 0.1 between hESCs and the differentiated cell types. The color code indicates the DNAme state distribution in the differentiated cell types, revealing that most regions reside in an IMR state after they lost DNAme (**left**). Genomic features (bottom) associated with DMRs losing DNAme in each of the differentiated cell types based on RefSeq gene annotation and *de novo* discovered promoters by RNA-Seq.

B. Venn diagram of identified DMRs that decrease their DNA methylation level between the three hESC-derived populations.

C. DNAme at the *POU3F1* locus (chr1:38,493,152-38,532,618). The heat map below shows the DNA methylation values of individual CpGs within the grey region. The average DNAme value for the entire highlighted region is shown on the right in red. CGIs are shown as green bars. Expression values (FPKM) are displayed on the right.

D. Promoters (left) and distal elements (right) that gain DNAme separated by the changes in FPKM at associated genes.

E. Chromatin state in differentiated cell types at regions that loose DNAme during differentiation.

See also Figure S4.

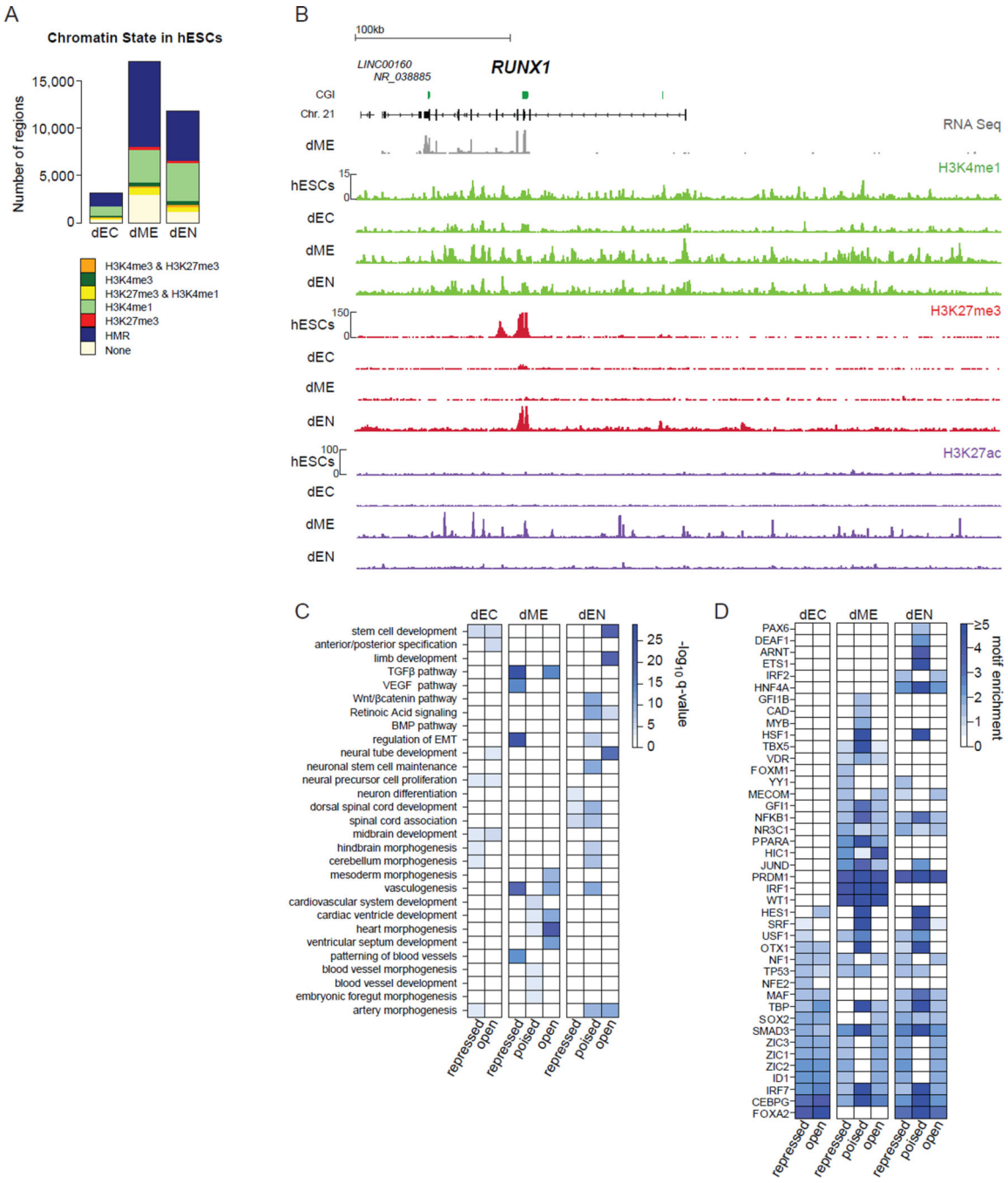

Figure 5. H3K27ac Dynamics Demarcate Novel Lineage Specific Gene Regulatory Elements

A. Number of regions and associated epigenetic state distribution in hESCs of regions that are transitioning to H3K27ac in the three populations.

B. Normalized ChIP-Seq tracks (H3K4me1, H3K27me3 and H3K27ac) for the *RUNX1* region (chr21:36,091,108-36,746,447) with corresponding RNA-Seq data in dME.

C. GO categories enriched in regions transitioning toH3K27ac in the cell type indicated on the right compared to hESCs as determined by GREAT analysis. Regions gaining H3K27ac were split up by state of origin in hESC into repressed (None, IMR, HMR, HK27me3),

poised (H3K4me1/H3K27me3) and open (H3K4me3/H3K27me3, H3K4me3, H3K4me1). Color code indicates multiple testing adjusted q-value of category enrichment.
D. TF motifs enriched in regions changing to H3K27ac in the cell type indicated on the right compared to hESCs. Color code indicates motif enrichment score incorporating total enrichment over background as well as differential expression of the corresponding transcription factor in the respective cell type. Regions were split up by state of origin in hESCs similar to panel **C**. For each region class, the eight highest-ranking motifs are shown.
See also Figure S5.

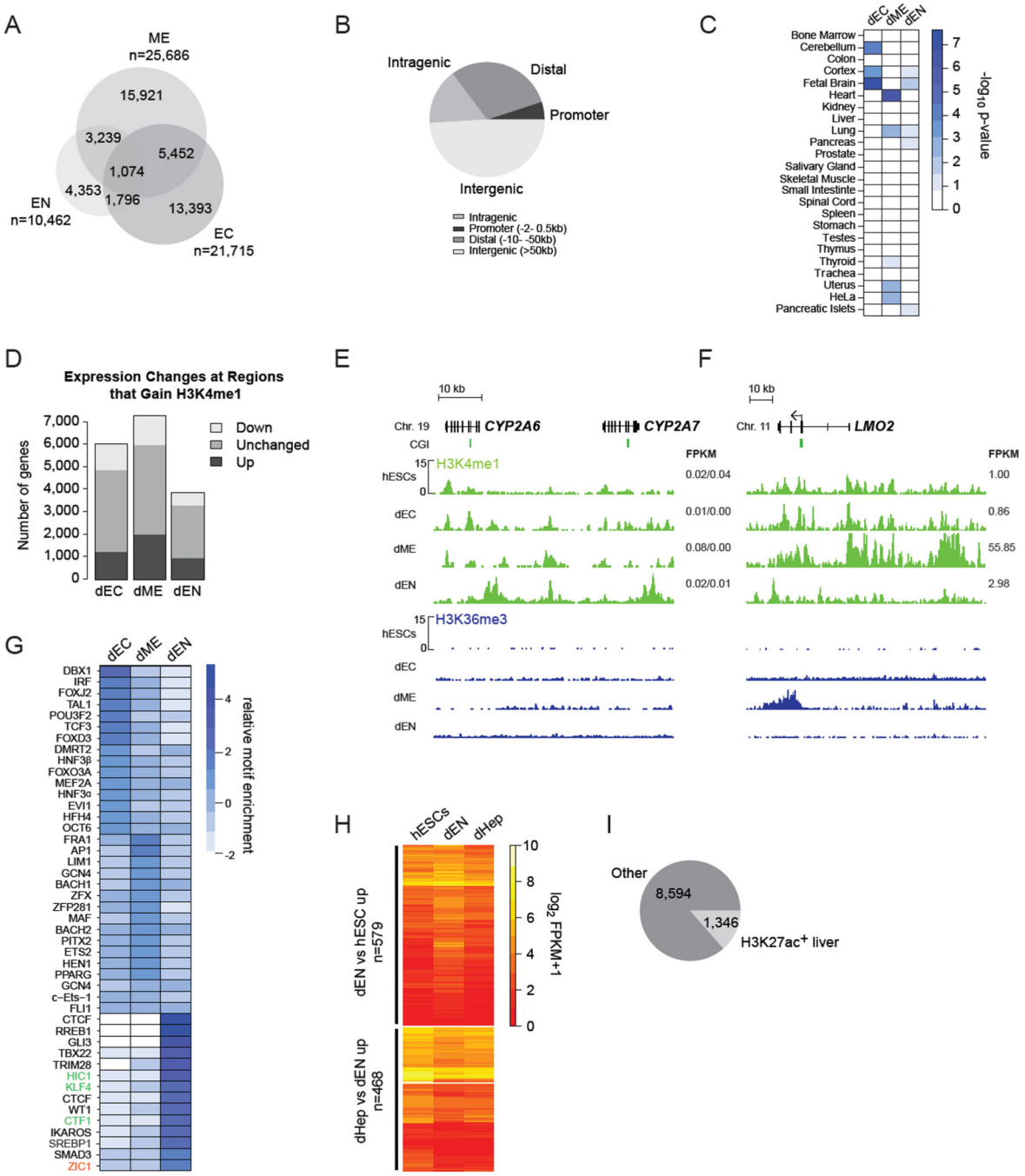

Figure 6. Characterization of H3K4me1 Dynamics at Putative Distal Regulatory Elements

A. Overlap of regions gaining H3K4me1 in the three differentiated populations relative to hESCs.

B. Genomic distribution of all regions gaining H3K4me1 compared to hESCs in at least one of the three differentiated populations.

C. Tissue signature enrichment levels of genes assigned to regions specifically gaining H3K4me1 in the differentiated populations indicated on the bottom. For tissue signature definitions, see Extended Experimental Procedures.

D. Number and distribution of gene expression changes of genes assigned to regions gaining H3K4me1 in the differentiated populations. Associated genes were classified as either being up/down- regulated or unchanged relative to hESCs.

E. Normalized ChIP-Seq tracks (H3K4me1 and H3K36me3) for the *LMO2* locus (Chr. 11:33,865,134-33,977,858). Read counts on y-axis are normalized 10 million reads for each cell type. CGIs are indicated in green.

F. Normalized ChIP-seq tracks (H3K4me1 and H3K36me3) for the *CYP2A6/CYP2A7* region (Chr19: 41,347,260-41,395,599). Read counts on y-axis are normalized 10 million reads for each cell type. CGIs are indicated in green.

G. Normalized motif enrichment scores for the top 15 motifs enriched in regions specifically transitioning to H3K4me1 in the differentiated cell type indicated on the bottom. Motif highlighted in red corresponds to a TF that is upregulated at the next stage (hepatoblast) of endoderm differentiation while motifs highlighted in green are specifically upregulated in dEN but downregulated at the dHep stage.

H. Gene expression levels of genes assigned to regions gaining H3K4me1 specifically in dEN compared to hESC and being upregulated in dEN but not hepatoblast (**top**). Gene expression levels of genes being upregulated between dEN and dHep (but not between hESC & dEN) and gaining H3K4me1 in dEN are shown on the bottom.

I. Fraction of regions changing to H3K4me1 in dEN and being enriched for H3K27ac in human liver (n=1,346).

See also Figure S6.

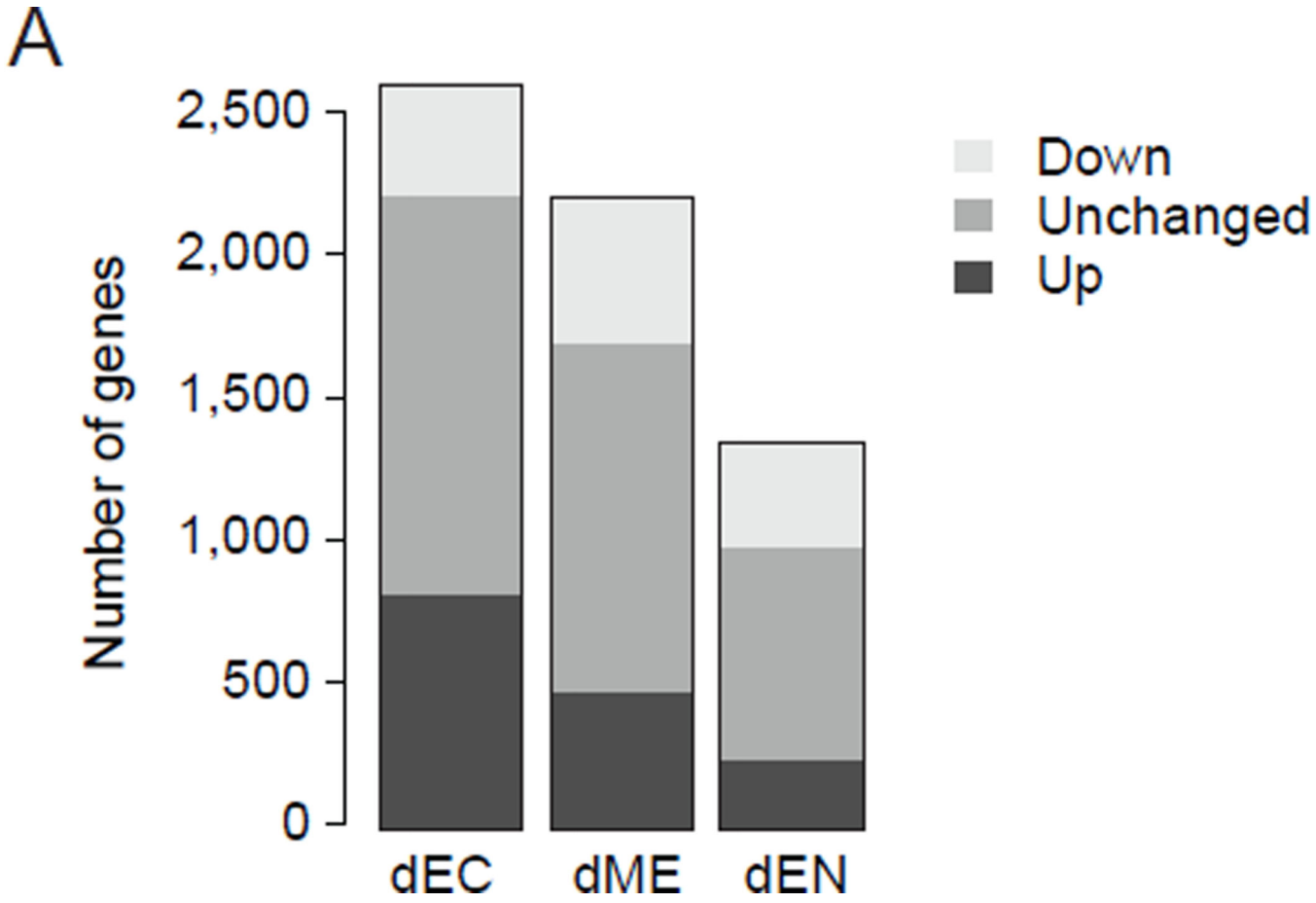

Figure 7. DNAme to H3K27me switch and FOXA2 Binding

A. Distribution of genomic features associated with region gaining H3K27me3 (n=22,643) upon differentiation to any of the three hESC derived cell types compared to hESC.

B. CpG content distribution of regions gaining H3K27me3 upon differentiation. For reference, the CpG content distribution of CpG islands is shown.

C. Epigenetic state distribution in hESC, dEC and dME of regions that gain H3K27me3 in the dEN population compared to hESC.

D. Binding profile of FOXA2 in dEN (n=357), OCT4 (n=32), SOX2 (n=12) and NANOG (n=124) in hESC across regions that gain H3K27me3 in dEN upon differentiation.

E. Composite plot of median normalized tag counts (RPKM) of regions bound by FOXA2 in dEN and gaining H3K27me3 in dEN compared to hESC (n=369).

F. Normalized H3K27me3 and H3K4me3 ChIP-seq tracks for hESCs, dEN and human adult liver tissue at the *ALB* locus (chr4:74,257,882-74,377,753). Black bars (**bottom**) indicate TF binding of OCT4, SOX2 or NANOG in hESCs. Read counts on y-axis are normalized to 10 million reads.

G. Distribution of methylation levels of regions bound by FOXA2 and gaining H3K27me3 in dEN. DNAme information is depicted for hESC and dEN WGBS datasets and two biological replicates of FOXA2 ChIP-Bisulfite experiments in dEN (n=369).

H. Gene expression profile of genes upregulated at the hepatoblast stage relative to dEN that are associated with regions bound by FOXA2 and gaining H3K27me3 in dEN (n=50).

I. Fraction of regions gaining H3K27me3 in dEN and being enriched for H3K27ac in human liver (n=192).

See also Figure S7.